Study Title	A Randomized, Double-Blinded, Placebo-Controlled Study for the Treatment of Ocular Chronic GVHD with Processed Amniotic Fluid (pAF) Drops (GVHD)
ClinicalTrials.gov ID (NCT Number)	NCT03298815
Principal Investigator (PI)	Catherine Lee, MD
Document (ICF, Protocol, SAP)	SAP
Update Date (Approval Date)	November 13, 2019

Statistical Analysis Plan

Protocol Title (Number): A Randomized, Double-Blinded, Placebo-Controlled Study for the Treatment of Ocular Chronic Graft-Versus-Host Disease (GVHD) with processed Amniotic Fluid (pAF) Drops (PROTOCOL NUMBER)

Protocol Version and Date: 1.00; October 16, 2019

SAP Author: Lawrence J Ccook, Ph.D.

SAP Version and Date: 1; November 13, 2019

SAP Version 1 Date: ORIGINAL DATE OF SAP APPROVAL

CONFIDENTIAL

Approvals:

Approved By:

Name	Title	Date
Name	Title	Date

Contents

Abbreviations

Abbreviation	Definition
DCC	Data Coordinating Center
CRF	Case Report Form
DSMB	Data and Safety Monitoring Board
ITT	Intent-To-Treat
PP	Per-Protocol
(S)AE	(Serious) Adverse Event
SAFETY	Safety Population
SAP	Statistical Analysis Plan

1 PREFACE

1.1 Purpose of SAP

This Statistical Analysis Plan (SAP) describes the planned analysis and reporting for the protocol: A Randomized, Double-Blinded, Placebo-Controlled Study for the Treatment of Ocular Chronic Graft-Versus-Host Disease (GVHD) with processed Amniotic Fluid (pAF) Drops .

This study is a double blind, randomized controlled trial of persons with ocular graph versus host disease. The goal of this pilot study is to collect preliminary data on the safety and efficacy of processed amniotic fluid to reduce pain resulting from dry eye.

The purpose of this SAP is to outline the planned analyses to be completed for the Ocular GVHD trial. The planned analyses identified in this SAP will be included in future study abstracts and manuscripts. Also, exploratory analyses not necessarily identified in this SAP may be performed. Any post hoc, or unplanned, analyses not explicitly identified in this SAP will be clearly identified as such in any published reports from this study.

1.2 Auxiliary/Other Documents

The following documents were reviewed in preparation of this SAP:

- Protocol: A Randomized, Double-Blinded, Placebo-Controlled Study for the Treatment of Ocular Chronic Graft-Versus-Host Disease (GVHD) with processed Amniotic Fluid (pAF) Drops .
- Case Report Forms (CRFs) for the Ocular GVHD protocol

The reader of this SAP is encouraged to read the protocol for details on the conduct of this study, and the operational aspects of clinical assessments.

It is possible that, due to updates or identification of errors in specific statistical software discussed in this SAP, the exact technical specifications for carrying out a given analysis may be modified. This is considered acceptable as long as the original, prespecified statistical analytic approach is completely followed in the revised technical specifications.

2 STUDY OBJECTIVES AND OUTCOMES

2.1 Study Objectives

2.1.1 Primary Objective(s)

The primary objectives of the Ocular GVHD trial are:

1. To determine the safety of pAF in patients with ocular chronic GVHD
2. To determine the clinical effects of pAF in ocular chronic graft versus host disease

2.1.2 Secondary Objective(s)

The secondary objectives of the Ocular GVHD trial are:

1. To determine the change in National Institutes of Health (NIH) Consensus Criteria (CC) ocular score of chronic GVHD at 30 ± 3 days, 60 ± 3 days, and 100 ± 3 days from baseline related to the administration of pAF
2. To determine change in Functional Assessment of Cancer Therapy-General (FACT G) Quality of Life (QOL) at 30 ± 3 days, 60 ± 3 days, and 100 ± 3 days related to the administration of pAF
3. To determine changes in visual acuity related to the administration of pAF
4. To determine the effects of pAF on the corneal surface
5. To determine the changes in dry eye symptoms using the grading provided by the International Dry Eye Workshop (DEWS) 2007 report
6. To determine the changes in patient reported pain level related to the administration of pAF using the 0-10 pain rating scale during the first 60 days of the study

2.2 Study Outcomes

2.2.1 Primary Outcome(s)

The primary outcome is the binary variable of complete or partial response to treatment (as defined by resolution or no worsening on the clinician- reported NIH CC eye score) and a decrease of at least one point on the DEWS.

2.2.2 Secondary Outcomes

The secondary outcomes are

1. Change in the NIH CC ocular score of chronic GVHD
2. Change in FACT G
3. Change in visual acuity
4. To determine the effects of pAF on the corneal surface
5. Change in DEWS
6. Changes in patient reported pain level

3 STUDY DESIGN AND METHODS

3.1 Overall Study Design

This is a randomized, double-blinded, placebo-controlled study of the efficacy of pAF in patients with hematologic malignancies who have undergone Allogeneic Hematopoietic Stem Cell Transplantation (HSCT) and are diagnosed with chronic GVHD of the eye. Patients will be allowed to continue the use of non-medicated lubricant eye drops (artificial tears) for ocular GVHD treatment prior to participating in this study. Patients are allowed to resume the use of eye drops after the 30 day treatment period at the discretion of the treating physician.

Participants will serve as their own control. A person will have one eye randomized to receive pAF and the other eye will receive saline drops.

3.2 Method of Treatment Assignment and Randomization

Upon acceptance into the study, the CTRM facility will be notified and the patient will be randomized as to which eye will receive the drug and which will receive the placebo. Randomization will be done using the MS Excel randomization method as developed by CTRM.

3.2.1 Delivery of Randomization and Emergency Backup

Based on the results of the randomization selection, CTRM will then select appropriate coded boxes. Four boxes containing 8-eyedropper bottles per box will be labeled with a yellow dot and stamped with an R for right eye. Each eyedropper bottle is labeled with a

yellow dot and stamped with an R. Four boxes containing 8-eyedropper bottles per box will be labeled with a blue dot and stamped with an L for left eye. Each eyedropper bottle is labeled with a blue dot and stamped with an L. The boxes will be labeled as Amniotic Fluid Eye Drops, each set (right and left) will contain a total of 32 eyedropper bottles of pAF and 32 eyedropper bottles of placebo.

3.3 Treatment Masking (Blinding)

How are handling treatment assignment in the database? We are going to need to know more than left vs. right. Do we have the treatment labels entered as a form? Why do I have so many questions?

The ocular GVHD study will be performed in in a double-blind fashion. All study personnel, including investigators and research coordinators will be blinded as to which eye is receiving the active treatment and which is receiving control.

Of necessity, biostatisticians involved in presenting interim analyses to the DSMB will be aware which subjects have received "Treatment A" and which have received "Treatment B", but they will not be aware of the identity of the two arms.

3.4 Study Intervention Compliance

Subjects will complete a daily diary indicating their use of the drops in the morning and evening. Participants will note each time there is an issue or the drops are used in a method different than described by the protocol.

4 STUDY SUBJECTS AND ANALYSIS POPULATIONS

4.1 Eligibility

Patients will be eligible for enrollment if they meet all of the following inclusion criteria:

1. Patients with chronic GVHD $\geq$ Grade II diagnosed within 3 years after hematopoietic stem cell transplant (HSCT) for any disease, with any graft, and any conditioning regimen with dry eye (patients may be using bilateral scleral lenses and/or bilateral punctal plugs at the time of accrual)
2. Patients who are 18 years of age or older.
3. Willing and able to provide informed consent.

Patients will be ineligible for enrollment if any of the following exclusion criteria are met:

- Patients who have any other reversible cause for dry eye at the time of accrual.
- More than 2 lines of therapy beyond corticosteroids with or without calcineurin inhibitors or sirolimus
- Relapsed malignancy after transplantation
- A difference in dryness between both eyes of more than 2 points of the grading provided by the International Dry Eye Workshop (DEWS) 2007 report
- Patients who are pregnant or plan to become pregnant while participating in the study.
- Patients who are not willing to discontinue the use of any eye drops, with the exception of non-medicated lubricant eye drops (artificial tears). All eye drops (excluding non-medicated lubricant eye drops) must be stopped at least seven days before treatment with pAF.
- Inability to comply with the investigational plan and visit schedule for any reason, in the judgement of the investigator.

4.2 Populations for Analyses

4.2.1 Screening Population

What do we want? All screened or all eligibile? The screening population (SCREEN) includes all patients who are screened for eligibility into the trial, regardless of randomization into the trial or treatment status. This population represents all patients who meet inclusion criteria outlined in the study protocol and who are screened. This population will be used for reporting of study flow per CONSORT guidelines.

4.2.2 Intention-to-Treat Population

The Intention-to-Treat (ITT) population includes all subjects who are randomized into the trial, regardless of adherence to the protocol, including, for example, subjects who receive no study drug. The ITT population will be used for the primary efficacy analyses in the study, as well as for main efficacy analyses of secondary outcomes. All analyses using the ITT population will be based on each subject's assigned treatment arm, regardless of treatment actually received.

4.2.3 Per-Protocol Efficacy Population

The Per-Protocol (PP) efficacy population includes all subjects in the ITT population who are verified to meet all study inclusion and exclusion criteria, who receive study drug according to their assigned study arm and used the drops as specified in the protocol. This population will be used to examine whether results seen in the ITT population are maintained in the population adhering to the protocol.

4.2.4 Safety Population

The safety population (SAFETY) includes all subjects who receive any study drug. Reporting of results based on this population will be summarized according to treatment received. This population will be used for analysis of adverse events and (in addition to ITT) to examine safety outcomes.

5 GENERAL ISSUES FOR STATISTICAL ANALYSES

5.1 Analysis Software

Analysis will be performed using SAS® Software version 9.4 or later whenever possible. Other software packages, including R and StatXact®, may be used for particular specialized procedures.

5.2 Methods for Withdrawals, Missing Data, and Outliers

Per the intention-to-treat principle, subjects who withdraw from the study or are lost to follow-up will have all available data used in the analysis. In the event that a substantial number of subjects are withdrawn or lost to follow-up, baseline characteristics and available information on hospital course will be reviewed and compared to subjects not withdrawn or lost, to assess empirically if these subjects differ from those remaining in the study for the scheduled treatment and follow-up time.
what do we do about participants who miss study visits?

5.3 Multiple Comparisons and Multiplicity

As there is a single primary endpoint for this study, adjustment for multiple comparisons will not be required for the primary analysis.

For the secondary efficacy endpoints, a Bonferroni-Holm stepdown test will be used for assessing significance, with a total alpha level of 0.05 for these two outcomes.

Safety outcomes for this study will be reported using unadjusted p- values for each individual comparison. However, all reports of these outcomes, to the DSMB and in published reports, will explicitly note that multiple safety outcomes have been evaluated. Formal multiplicity-adjusted significance assessments for these outcomes will be performed upon request of the DSMB or other reviewers. The Bonferroni-Holm procedure will be used for such assessments.

5.4 Planned Subgroups, Interactions, and Covariates

There are no planned subgroup analyses.

5.5 Derived and Computed Variables

The primary outcome for the study, response vs. non-response will be defined according to the following table.

Table 1: Response Table

Response Type	Definitions
Complete Response	Resolution of all manifestations of ocular GVHD 1. NIH CC eye score = 0 2. DEWS = 1
Partial Response (PR)	1. No worsening in NIH CC eye score 2. At least one point decrease in DEWS
Mixed Response	1a. Any decrease in the NIH CC eye score 1b No improvement in DEWS. or 2a. Decrease in DEWS 2b. Worsening in the NIH CC eye score
No Response (NR)	No change in both the NIH CC score and the DEWS
Progressive Disease (PD)	Worsening on DEWS irrespective of NIH CC eye score

An eye will be coded as responding if it is categorized as CR or PR. All other situations, including MR, NR, and PD will be considered as not responding.

5.6 Independent Review

All statistical analyses for primary reporting of trial results will be independently verified through dual programming. Two statisticians will each program all datasets and analyses

for the DSMB reports and the primary manuscript(s) and the results will be compared. This process will begin at the analysis design stage and will continue through writing of abstracts and manuscripts.

6 PLANNED ANALYSES

6.1 Description of Subject Characteristics

Publication of the primary results will include reporting of key baseline characteristics overall. These will include, but are not limited to

- gender
- race
- age
- baseline FACT-G
- baseline NIH CC eye score for each eye
- baseline DEWS for each eye
- baseline patient reported pain level for each eye
- baseline visual acuity for each eye

6.2 Primary Outcome Analysis

6.2.1 Response Table

The outcome is the combination of two measures, the NIH CC criteria and DEWS. Each will be summarized at 30, 60, and 100 days.

The NIH CC is a four level ordinal score ranging from 0 to to 3 with 0 being asymptomatic and 3 being the worst possible. Treatment and control eyes will be summarized at baseline and each visit using counts and percentages. At each visit we will summarize the number and percent of treatment and control eyes that have improved, stayed the same, and progressed compared to baseline. An improvement will be defined as movement to a lower score and symptom progression will be defined as movement to a higher score.

The DEWS is a four level ordinal score ranging from 1 to 4 with 1 being asymptomatic to mild and 4 being the severe. Similar to the NIH CC score, treatment and control eyes will

be summarized at baseline and each visit using counts and percentages. At each visit we will summarize the number and percent of treatment and control eyes that have improved, stayed the same, and progressed compared to baseline. An improvement will be defined as movement to a lower score and symptom progression will be defined as movement to a higher score.

Analysis of the primary endpoint will be coded as each eye having responded or not responded at the 30 day visit. Because a participant's eyes are more likely to be similar compared to eyes from two different participants we will conduct a paired analysis using McNemar's test, as implemented in PROC FREQ, by constructing the following table where A is the number of participants who had both eyes respond; B is the number of participants who responded in the active eye but not in the control eye; C is the number of participants who did not respond in the active eye but responded in the control eye; and D is the number of participants who did not respond in either eye. pAF will be deemed to be effective if B is significantly greater than its expected value. If the observed table from the trial is sparse we will utilize the exact version of the McNemar's test.

Table 2: Outcome Table

		Control Eye	
		Responding	Not Responding
Active Eye	Responding	A	B
	Not Responding	C	D

6.3 Secondary Outcome(s) Analyses

6.3.1 NIH Consensus Criteria Ocular Score of Chronic GVHD

The NIH CC will be summarized as above. At each follow-up visit, 30 days, 60 days, and 100 days, the percent of treated eyes that have improved and progresses will be summarized using 95% confidence intervals. To test for a change within participants a sign test will be conducted. An improvement in an eye will be coded as a +1, symptom progression in an eye will be coded as -1, and no change coded as 0. A score will be created for each participant by subtracting the control eye from the treatment eye. Given the sample size it is likely we will use an exact binomial p-value.

6.3.2 Dry Eye Severity Grading Scheme (DEWS)

The DEWS will be summarized as above. At each follow-up visit, 30 days, 60 days, and 100 days, the percent of treated eyes that have improved and progressed will be summarized using 95% confidence intervals. To test for a change within participants a sign test will be conducted. An improvement in an eye will be coded as a +1, symptom progression in an eye will be coded as -1, and no change coded as 0. A score will be created for each participant

by subtracting the control eye from the treatment eye. Given the sample size it is likely we will use an exact binomial p-value.

6.3.3 Functional Assessment of Cancer Therapy - General (FACT G)

The FACT-G will be summarized at baseline at the 30, 60, and 90 day follow-up visits using medians and interquartile ranges. The overall FACT-G is calculated by scoring the individual subsections and then deriving an overall score (we need the manual to do this). FACT-G will not be normalized to a 0 - 100 range. A Wilcoxon signed-rank test will be calculated to compare change from baseline at each visit by subtracting the follow-up visit score from the baseline score.

6.3.4 Visual Acuity

For this outcome it is important to know the percent of subjects with changes, improved or decreased, and the magnitude of the change. To summarize improvement, decrease, or no change at each follow-up visit each eye will be scored as a +1 for an increase in visual acuity, -1 for a decrease in visual acuity, and 0 for no change. Counts and percents will be used to summarize change in visual acuity for both the treatment and control eye. As with the NIH CC and DEWS, the control eye change will be subtracted from the the treatment eye change and a sign test will be conducted.

To estimate a the magnitude of change in acuity, Snellen Chart scores will be converted to LogMAR value. The change in acuity at follow-up for each eye will then be calculated as baseline LogMAR - follow-up LogMAR. The treatment eye change - the control eye change will then be used to conduct a Wilcoxon signed-test.

6.3.5 Effect of pAF on the Corneal Surface

Changes in the surface of the cornea will be measured using epithelial staining and graded on the following scale: Trace, 1+, 2+, 3+, 4+, Confluent. An improvement at follow-up will be considered to be movement to a lower grade. The treatment and control eyes will be scored at each visit with a +1 for improvement, -1 for decline, and 0 for no change. The number of treatment and control eyes improving and declining at each visit will summarized using 95% confidence intervals. To test for changes within patients the control eye will be subtracted from the treatment eye and a sign test will be conducted.

6.3.6 Pain

Change in pain will be calculated from the patient diaries. A first level summary of pain will be to create median pain scores by eye for each week. Median trace plots and side-by-side box

plots will then be used to display the change in pain over time for treatment and control eyes.

The individual pain scores for each eye will be used to estimate a regression slope for each eye with pain as the outcome and day of treatment as the explanatory variable. The estimated slope of the control eye will be subtracted from the slope of the treatment eye and a Wilcoxon signed-rank test will be conducted.

6.4 Safety Analyses

The proportion of subjects who experience adverse events and serious adverse events will be summarized descriptively; no inferential tests will be performed.

7 SAMPLE SIZE DETERMINATION

Analysis of the primary endpoint will be coded as each having responded or not responded at the 30 day visit. Because a participants eyes are more likely to be similar compared to eyes from two different participants we will conduct a paired analysis using McNemars test. Assuming a significance level of 0.05, and a response rate in the control eye of 5%, 15 participants will provide 80% power to detect a difference if the response in the treated eye is 60%. With 19 participants we will have 80% power to detect a response rate of 50%. Due to the small sample size it is likely an exact p-value will be calculated.

8 References